



Министерство науки и высшего образования Российской Федерации  
федеральное государственное автономное  
образовательное учреждение высшего образования  
«Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа Инженерная школа информационных технологий и робототехники  
Направление подготовки 09.03.01 Информатика и вычислительная техника  
Отделение школы (НОЦ) Отделение информационных технологий

### БАКАЛАВРСКАЯ РАБОТА

Тема работы
<b>Разработка Web-модуля для парсинга данных с сайтов торговых площадок</b>

УДК 004.624:004.774:339.371.5

Студент

Группа	ФИО	Подпись	Дата
8В5Б	Сутягина Анна Андреевна		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Саврасов Федор Витальевич	К.Т.Н.		

### КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент ОСГН	Шулинина Юлия Игоревна			

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент ООТД	Мезенцева Ирина Леонидовна			

### ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Погребной Александр Владимирович	К.Т.Н.		

## ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ООП

Код результата	Результат обучения (выпускник должен быть готов)
P1	Применять базовые и специальные естественнонаучные и математические знания в области информатики и вычислительной техники, достаточные для комплексной инженерной деятельности
P2	Применять базовые и специальные знания в области современных информационных технологий для решения инженерных задач
P3	Ставить и решать задачи комплексного анализа, связанные с созданием аппаратно-программных средств информационных и автоматизированных систем, с использованием базовых и специальных знаний, современных аналитических методов и моделей
P4	Разрабатывать аппаратные и программные средства (системы, устройства, блоки, программы, базы данных и т.п.) в соответствии с техническим заданием и с использованием средств автоматизации проектирования
P5	Проводить теоретические и экспериментальные исследования в области создания аппаратных и программных средств информационных и автоматизированных систем, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретацию полученных данных
P6	Внедрять, эксплуатировать и обслуживать современные программно-аппаратные комплексы, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья, безопасность труда, выполнять требования по защите окружающей среды
P7	Использовать базовые и специальные знания в области проектного менеджмента для ведения комплексной инженерной деятельности
P8	Владеть иностранным языком на уровне, позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты комплексной инженерной деятельности
P9	Эффективно работать индивидуально и в качестве члена группы, состоящей из специалистов различных направлений и квалификаций, демонстрировать ответственность за результаты работы и готовность следовать корпоративной культуре организации
P10	Демонстрировать знания правовых, социальных, экономических и культурных аспектов комплексной инженерной деятельности
P11	Демонстрировать способность к самостоятельному обучению в течение всей жизни и непрерывному самосовершенствованию в инженерной профессии

Министерство науки и высшего образования Российской Федерации  
федеральное государственное автономное  
образовательное учреждение высшего образования  
«Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа Инженерная школа информационных технологий и робототехники  
Направление подготовки 09.03.01 Информатика и вычислительная техника  
Отделение школы (НОЦ) Отделение информационных технологий

УТВЕРЖДАЮ:  
Руководитель ООП

\_\_\_\_\_  
(Подпись)      (Дата)      (Ф.И.О.)

### ЗАДАНИЕ на выполнение выпускной квалификационной работы

В форме:

Бакалаврской работы
(бакалаврской работы, дипломного проекта/работы, магистерской диссертации)

Студенту:

Группа	ФИО
8B5Б	Сутягиной Анне Андреевне

Тема работы:

<b>Разработка Web-модуля для парсинга данных с сайтов торговых площадок</b>	
Утверждена приказом директора (дата, номер)	От 15.02.2019 №1217/с

Срок сдачи студентом выполненной работы:

--	--

#### ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<b>Исходные данные к работе</b>	Разрабатываемый модуль должен выгружать необходимую информацию в базу данных с сайтов различных торговых площадок. При разработке необходимо учесть, что модуль должен быть ориентирован на дальнейшее включение в системы интеллектуального анализа данных.
<b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b>	Обзор литературы, обзор аналогичных проектов, выработка требований, проектирование архитектуры модуля, выбор инструментов для работы (языки программирования, библиотеки)

<b>Перечень графического материала</b>		Схема базы данных, диаграмма архитектуры модуля
<b>Консультанты по разделам выпускной квалификационной работы</b>		
<b>Раздел</b>	<b>Консультант</b>	
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Шулинина Юлия Игоревна	
Социальная ответственность	Мезенцева Ирина Леонидовна	

<b>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</b>	
---	--

**Задание выдал руководитель:**

<b>Должность</b>	<b>ФИО</b>	<b>Ученая степень, звание</b>	<b>Подпись</b>	<b>Дата</b>
Доцент ОИТ	Саврасов Федор Витальевич	к.т.н.		

**Задание принял к исполнению студент:**

<b>Группа</b>	<b>ФИО</b>	<b>Подпись</b>	<b>Дата</b>
8В5Б	Сутягина Анна Андреевна		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА  
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И  
РЕСУРСОБЕРЕЖЕНИЕ»**

Студенту:

<b>Группа</b>	<b>ФИО</b>
8В5Б	Сутягиной Анне Андреевне

<b>Школа</b>	<b>ИШИТР</b>	<b>Отделение школы (НОЦ)</b>	<b>ОИТ</b>
<b>Уровень образования</b>	бакалавриат	<b>Направление/специальность</b>	09.03.01 Информатика и вычислительная техника

**Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:**

1. <i>Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	Оклад инженера – 21760 руб. Оклад руководителя – 33664 руб.
2. <i>Нормы и нормативы расходования ресурсов</i>	Премияльный коэффициент 30%; Коэффициент доплат и надбавок 35%; Районный коэффициент 30%; Коэффициент дополнительной заработной платы 12%; Накладные расходы 16%.
3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	Коэффициент отчислений на уплату во внебюджетные фонды 30%

**Перечень вопросов, подлежащих исследованию, проектированию и разработке:**

1. <i>Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения</i>	-Анализ конкурентных технических решений
2. <i>Планирование и формирование бюджета научных исследований</i>	Формирование плана и графика разработки: - определение структуры работ; - определение трудоемкости работ; - разработка графика Ганта. Формирование бюджета затрат на научное исследование: - материальные затраты; - затраты на специальное оборудование; - заработная плата (основная и дополнительная); - отчисления на социальные цели; - накладные расходы.
3. <i>Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования</i>	- Определение потенциального эффекта исследования

**Перечень графического материала (с точным указанием обязательных чертежей):**

1. <i>Оценочная карта конкурентных технических решений</i>
2. <i>Матрица SWOT</i>
3. <i>График Ганта</i>
4. <i>Расчет бюджета затрат</i>

**Дата выдачи задания для раздела по линейному графику**

**Задание выдал консультант:**

<b>Должность</b>	<b>ФИО</b>	<b>Ученая степень, звание</b>	<b>Подпись</b>	<b>Дата</b>
Ассистент ОСГН ШБИП	Шулинина Ю.И.			

**Задание принял к исполнению студент:**

<b>Группа</b>	<b>ФИО</b>	<b>Подпись</b>	<b>Дата</b>
8В5Б	Сутягина А.А.		

## ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8В5Б	Сутягиной Анне Андреевне

Школа	ИШИТР	Отделение (НОЦ)	ОИТ
Уровень образования	бакалавриат	Направление/специальность	09.03.01 Информатика и вычислительная техника

Тема ВКР:

«Разработка Web-модуля для парсинга данных с сайтов торговых площадок»	
<b>Исходные данные к разделу «Социальная ответственность»:</b>	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Объектом исследования является разработка Web-модуля для извлечения данных о товарах и отзывах с Интернет-площадок. Данный модуль создается для дальнейшего включения в системы интеллектуального анализа данных.
<b>Перечень вопросов, подлежащих исследованию, проектированию и разработке:</b>	
<b>1. Правовые и организационные вопросы обеспечения безопасности:</b> <ul style="list-style-type: none"> <li>– специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства;</li> <li>– организационные мероприятия при компоновке рабочей зоны.</li> </ul>	Анализ трудового кодекса Российской Федерации (ТК РФ) и санитарно-эпидемиологических правил и нормативов (СанПиН 2.2.2/2.4.1340-03).
<b>2. Производственная безопасность:</b> 2.1. Анализ вредных и опасных факторов при проведении исследований и при использовании объекта исследования и обоснование мероприятий по защите исследователя	Возможны следующие вредные и опасные факторы: <ul style="list-style-type: none"> <li>– электромагнитные поля;</li> <li>– электростатические поля;</li> <li>– шум;</li> <li>– нарушения микроклимата в помещении;</li> <li>– нарушения освещенности рабочей зоны;</li> <li>– психофизические факторы.</li> </ul>
<b>3. Экологическая безопасность:</b>	Влияние на окружающую среду проявляется в: <ul style="list-style-type: none"> <li>– увеличении энергопотребления;</li> <li>– отсутствии утилизации ПЭВМ;</li> <li>– отсутствие утилизации макулатуры.</li> </ul>
<b>4. Безопасность в чрезвычайных ситуациях:</b>	Чрезвычайная ситуация техногенного характера для рассматриваемого случая – пожар.

Дата выдачи задания для раздела по линейному графику	
--	--

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
ассистент ООТД	Мезенцева И.Л.			

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8В5Б	Сутягина А.А.		

Министерство науки и высшего образования Российской Федерации  
 федеральное государственное автономное  
 образовательное учреждение высшего образования  
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа Инженерная школа информационных технологий и робототехники  
 Направление подготовки (специальность) 09.03.01 Информатика и вычислительная техника

Уровень образования бакалавр

Отделение школы (НОЦ) Отделение информационных технологий

Период выполнения \_\_\_\_\_ (осенний / весенний семестр 2018 /2019 учебного года)

Форма представления работы:

Бакалаврская работа
---------------------

(бакалаврская работа, дипломный проект/работа, магистерская диссертация)

### КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	
--	--

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
20.05.2019	Написание основной части	60
27.05.2019	Написание части финансового менеджмента, ресурсоэффективности и ресурсосбережения	15
01.06.2019	Написание части социальной ответственности	15
19.06.2019	Защита работы	10

#### СОСТАВИЛ:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Саврасов Федор Витальевич	к.т.н.		

#### СОГЛАСОВАНО:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Погребной Александр Владимирович	к.т.н.		

## **Реферат**

Выпускная квалификационная работа содержит 58 страниц, 9 рисунков, 13 таблиц, 19 литературных источников.

Ключевые слова: веб-разработка, веб-сервер, парсинг данных, извлечение данных, базы данных, синтаксический анализ, Интернет-торговля.

Объектом исследования является извлечение данных о товарах и отзывах с сайтов торговых площадок.

Предмет исследования – веб-модуль для извлечения информации с сайтов в базу данных.

Цель работы – проектирование и разработка веб-модуля для извлечения данных с возможностью его дальнейшего включения в системы интеллектуального анализа данных. Данная цель обусловлена необходимостью иметь структурированные и подготовленные для анализа большие объемы данных. Проблема сложности анализа информации человеком без подобных систем актуальна для пользователей Интернет-магазинов с огромным количеством отзывов на каждый товар.

В ходе работы были выработаны требования к разрабатываемому модулю, спроектирована архитектура и разработана первая рабочая версия модуля для торговой площадки Amazon, а также определено направление дальнейшего развития модуля.



## Определения

В данной работе используются следующие термины:

- парсинг: синтаксический анализ данных, применяемый для извлечения необходимой информации с веб-ресурсов;
- база данных (БД): структура взаимосвязанных, хранящихся вместе данных, организованная таким образом, что обеспечивается безызбыточное хранение данных, их комплексное использование и независимость программ обработки от физической структуры;
- система управления базами данных (СУБД): совокупность программных средств для создания баз данных и управления ими;
- HTML: язык разметки веб-ресурсов;
- CSS: язык форматирования, применяемый к элементам документов;
- XML: расширяемый язык разметки документов;
- регулярные выражения: шаблоны, используемые для сопоставления последовательностей символов в строке;
- DOM: объектная модель документа;
- XPath: язык запросов к элементам XML-документа;
- URL: единообразный указатель местонахождения ресурса;
- API: программный интерфейс приложения;
- HTTP: протокол передачи данных, основанный на технологии «клиент-сервер».

## Оглавление

Введение.....	12
ОСНОВНАЯ ЧАСТЬ.....	13
1 Литературный обзор.....	13
1.1 Парсинг веб-ресурсов.....	13
1.2 Обзор различных методов парсинга.....	14
1.3 Обзор аналогов.....	15
2 Техническое задание.....	17
2.1 Основные задачи и цели.....	17
2.2 Требования к модулю.....	17
3 Построение архитектуры приложения.....	18
4 Программная реализация модуля.....	21
4.1 Язык программирования.....	21
4.2 Веб-сервер.....	22
4.3 HTTP-запросы.....	22
4.4 Выгрузка необходимых данных.....	23
4.5 Проектирование базы данных.....	24
ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ.....	26
1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения.....	26
1.1 Потенциальные потребители результатов исследования.....	26
1.2 Анализ конкурентных технических решений.....	28
1.3 SWOT-анализ.....	31
2 Планирование научно-исследовательских работ.....	32
2.1 Структура работ в рамках научного исследования.....	32
2.2 Определение трудоемкости выполнения работ.....	33
2.3 Разработка графика проведения научного исследования.....	34
2.4 Бюджет научно-технического исследования.....	37
2.4.1 Расчет материальных затрат научно-технического исследования.....	37
2.4.2 Расчет затрат на специальное оборудование для научных (экспериментальных) целей.....	37

2.4.3	Основная заработная плата исполнителей темы.....	38
2.4.4	Дополнительная заработная плата исполнителей темы.....	40
2.4.5	Отчисления во внебюджетные фонды (страховые отчисления) ...	41
2.4.6	Накладные расходы.....	41
2.4.7	Формирование бюджета затрат научно-исследовательского проекта .....	42
3	Определение потенциального эффекта исследования .....	42
	СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ .....	44
1	Правовые и организационные вопросы обеспечения безопасности. ....	44
1.1	Специальные (характерные для проектируемой рабочей зоны) правовые нормы трудового законодательства. ....	44
1.2	Организационные мероприятия при компоновке рабочей зоны. ....	45
2	Профессиональная социальная безопасность. ....	46
2.1	Анализ вредных и опасных факторов при проведении исследований и при использовании объекта исследования и обоснование мероприятий по защите исследователя. ....	46
3	Экологическая безопасность. ....	53
3.1	Анализ влияния процесса исследования и объекта исследования на окружающую среду.....	53
3.2	Обоснование мероприятий по защите окружающей среды. ....	53
4	Безопасность в чрезвычайных ситуациях. ....	54
4.1	Анализ вероятных ЧС, которые могут возникнуть на рабочем месте при проведении исследований или дальнейшем использовании объекта исследований. ....	54
4.2	Обоснование мероприятий по предотвращению ЧС и разработка порядка действия в случае возникновения ЧС. ....	55
	Выводы .....	55
	Заключение .....	56
	Список литературы .....	57

## **Введение**

Использование возможностей Интернета в сфере торговли имеет множество преимуществ для потребителей, например, делает возможным рассмотреть весь спектр интересующих товаров дома, однако, на совершение покупки, особенно в Интернет-магазине, большое влияние оказывают отзывы других покупателей. Количество отзывов на товары с сайтов современных торговых площадок требует огромного количества времени для анализа человеком, в большинстве случаев найти интересующую информацию среди всех отзывов таким образом затруднительно.

Современным решением таких проблем является использование систем интеллектуального анализа больших данных. Однако для таких систем, прежде всего, нужны данные в структурированном виде. Поэтому при разработке систем анализа было бы гораздо удобнее использовать уже готовое решение для поиска и подготовки всех необходимых данных. Таким решением будет разрабатываемый в данной работе модуль.

Создание такого модуля предоставляет возможность получения всей необходимой для последующего анализа информации в удобном виде. При этом должна быть возможность легко модифицировать модуль, например, добавлять новые площадки для парсинга, изменять наборы извлекаемых данных, и т.д, поскольку данная работа ориентирована на включение в системы анализа данных.

## **ОСНОВНАЯ ЧАСТЬ**

### **1 Литературный обзор**

#### **1.1 Парсинг веб-ресурсов**

У современного человека часто возникает необходимость получения огромного количества информации с веб-ресурсов и иногда на поиск нужной информации уходит гораздо больше запланированного времени, к тому же, анализируя большие объемы информации, всегда есть вероятность упустить что-то важное.

В данном случае удобно использовать автоматизированный поиск информации, что возможно сделать либо с помощью API, либо с помощью парсинга сайта. Однако, использование API возможно только в том, случае, если его предоставляет веб-ресурс, поэтому подробнее рассмотрим процесс парсинга веб-ресурсов.

Парсинг — это синтаксический анализ веб-документа и извлечение необходимой информации. Эта методика преимущественно фокусируется на преобразовании неструктурированных данных в Интернете в структурированные данные: базы данных или электронные таблицы. Под парсером же подразумевается скрипт, используемый для автоматического сбора информации со страниц сайта и предоставления ее пользователю в структурированном виде. Парсер помогает выделять и импортировать повторяющиеся данные автоматически, существенно экономя время и предупреждая возможные ошибки копирования этой информации вручную. Парсинг сайта включает в себя доступ к Интернету напрямую через HTTP или через веб-браузер.

Работа парсера происходит следующим образом:

- 1) получение URL ресурса;
- 2) загрузка содержимого html-страницы;

3) поиск и извлечение всей необходимой информации.

## **1.2 Обзор различных методов парсинга**

Для извлечения данных в работе использована библиотека QueryList. Среди альтернативных вариантов были подробно рассмотрены использование регулярных выражений, модулей DOM и XPath, библиотеки Simple HTML DOM.

Регулярные выражения являются инструментом для извлечения информации с помощью шаблонов. В случае использования регулярных выражений переход на новую торговую площадку сопровождался бы созданием большого количества новых сложных регулярных выражений. Следовательно, чем больше торговых площадок бы охватывал модуль, тем большее количество регулярных выражений пришлось бы хранить и даже при незначительном изменении кода веб-ресурса регулярные выражения тоже приходилось бы корректировать. К тому же при больших объемах данных работа регулярных выражений занимает много времени.

Согласно DOM-модели, документ является иерархией («деревом»). Каждый HTML-тег образует узел дерева с типом «элемент». Вложенные в него теги становятся дочерними узлами. Для представления текста создаются узлы с типом «текст». Используя данный метод необходимые данные можно получить по идентификатору, имени, другим атрибутам элемента дерева или же с помощью уникального пути, спускаясь вниз по дереву. Однако путь к элементу может оказаться слишком сложным и его тоже необходимо менять при изменении структуры веб-ресурса.

XPath — это язык запросов к элементам XML- или XHTML-документа. Чтобы получить интересующие данные, необходимо создать запрос, описывающий эти данные.

Использование DOM и XPath уступает выбранной библиотеке отсутствием возможности работы с невалидным html-кодом.

Библиотеки QueryList и Simple HTML DOM, помимо отсутствия вышеупомянутых минусов, имеет удобные для поиска элементов HTML-кода возможности, например, обращение к HTML-элементам с помощью CSS селекторов, что значительно упрощает парсинг новых торговых площадок. Однако, производительность библиотеки QueryList выше, чем у Simple HTML DOM.

### **1.3 Обзор аналогов**

Среди имеющихся на данный момент средств извлечения данных с веб-ресурсов рассмотрим 2 наиболее близких конкурентных решения – программу для парсинга различных сайтов Datacol и сервис сбора и агрегации отзывов Mneniya.pro.

Парсер Datacol работает с регулярными выражениями и с XPath-запросами, имеет настройки для парсинга большого количества площадок. Интерфейс программы позволяет пользователю самостоятельно настроить парсер. Однако, для вышеуказанных целей данный продукт обладает такими недостатками, как закрытый исходный код, ручная настройка парсера и ограниченность бесплатной версии.

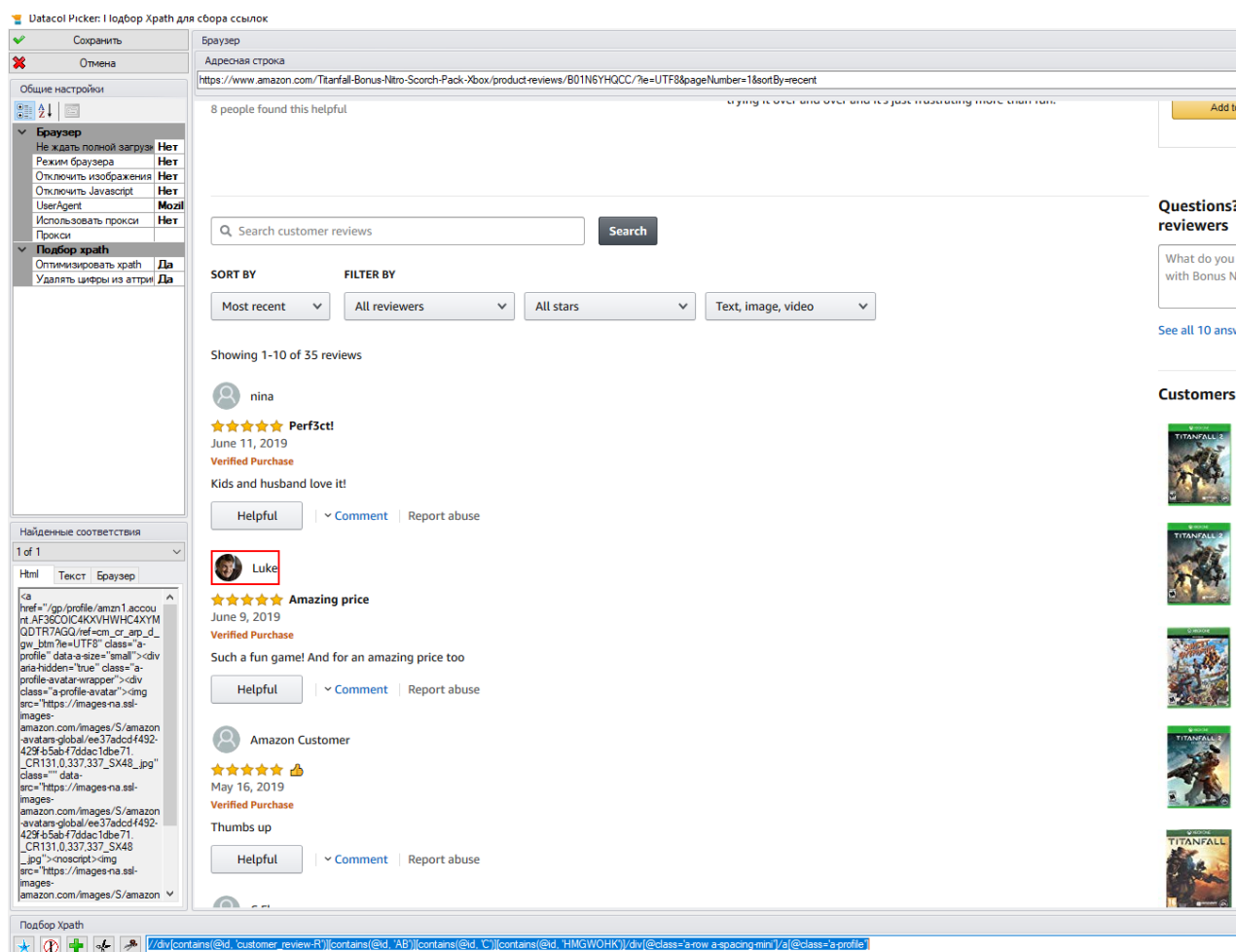


Рисунок 1 – Интерфейс настройки XPath-выражений парсера Datacol

Сервис Mneniya.pro также является платным, не имеет открытого исходного кода и помимо этого ориентирован на несколько другие задачи – вывод отзывов на сайт клиента.

Большинство аналогичных решений являются парсерами с графическим интерфейсом, являющиеся отдельным продуктом, в нашем же случае необходимо решение для легкого включения в более сложные системы, без лишнего функционала.

Разрабатываемый модуль будет выполнен более низкоуровневым для большей гибкости при включении в системы интеллектуального анализа и с открытым исходным кодом, доступным для модификации и расширения.



## **2 Техническое задание**

### **2.1 Основные задачи и цели**

Целью данной работы является разработка веб-модуля для извлечения данных о товарах и отзывах с сайтов торговых площадок в базу данных.

При получении ссылки на товар модуль должен будет выполнять следующие задачи:

- 1) Проверять правильность введения пользователем ссылки;
- 2) Определять торговую площадку и в соответствии с ней устанавливать параметры парсинга;
- 3) Запускать парсер;
- 4) Проверять созданную базу данных на наличие записей, соответствующих данному товару и отзывам;
- 5) Обновлять базу данных в случае обнаружения новых данных.

### **2.2 Требования к модулю**

Перед проектированием архитектуры модуля необходимо сформировать все предъявляемые требования:

- 1) В качестве входных данных модуль должен принимать полные рабочие ссылки на страницы с товарами, интересующими пользователя;
- 2) Модуль должен извлекать данные о товарах и отзывах с веб-ресурсов;
- 3) Все найденные данные должны сверяться с имеющимися в таблицах базы данных, в случае наличия новых база данных должна обновляться.
- 4) Модуль должен легко настраиваться на различные торговые площадки.

- 5) Исходный код разработанного модуля должен быть открытым и легким для понимания;
- 6) Модуль должен работать даже с невалидным HTML-кодом;
- 7) Модуль должен иметь высокую производительность даже в случае больших объемов данных;
- 8) Необходимы меры предупреждения блокировки IP-адреса;
- 9) Необходима ориентированность на включение в другие проекты (системы интеллектуального анализа данных);
- 10) Модуль должен быть кроссплатформенным.

### **3 Построение архитектуры приложения**

На рисунках 2 и 3 с помощью блок-схем представлена работа модуля.

В начале работы модуля происходит подключение библиотеки QueryList для создания парсеров, затем подключение к созданной БД, после чего модуль ожидает ввода пользователем ссылки на интересующую его страницу. Обязательным пунктом является проверка введенного адреса на валидность, используя шаблон, включающий все обязательные части ссылки на веб-ресурс. После успешной проверки ссылки с помощью функции языка PHP `parse_url` определяем торговую площадку для парсинга, в соответствии с которой выбираются настройки для парсеров товаров и отзывов.

Выгрузка данных начинается с отправки HTTP-запроса браузером к серверу торговой площадки с установленными заголовками для загрузки требуемой страницы. На данном этапе также принимаются меры для предупреждения блокировки IP-адреса. Модуль подключает внутренние файлы со списками различных User-Agent и Proxy строк и устанавливает в соответствующие заголовки случайные строки из этих файлов.

На запрос браузера сервер выдает ответ с кодом состояния, помогающим понять результат запроса. Успешная загрузка страницы HTTP-ответ будет определяться получением в HTTP-ответе кода 200, однако, на

данном этапе есть вероятность получения в HTTP-ответе кода с ошибкой. Наиболее распространённые коды состояний приведены в таблице 1. В соответствии с полученным кодом состояния пользователю будет выведено соответствующее сообщение с пояснениями, достаточными для понимания причины ошибки.

Таблица 1 – Коды состояний HTTP

Код состояния	Пояснение
200 OK	Успешный запрос
404 Not found	Запрос к несуществующему ресурсу
500 Internal server error	Наличие внутренней ошибки сервера
401 Unauthorized	Для запроса необходима аутентификация
403 Forbidden	Доступ к ресурсу запрещен
429 Too Many Requests	Запрос отклоняется из-за ограничения скорости

В случае успешного запроса модуль загружает страницу требуемого ресурса и с помощью установленных параметров для парсинга считывает необходимую информацию о товаре. Далее БД проверяется на наличие полученной информации в таблице item и при необходимости обновляется.

В цикле происходит выгрузка отзывов с текущей страницы и формирование URL следующей.

С появлением большого количества парсеров стала развиваться защита веб-ресурсов от парсинга, основанная на том, что работа парсера заметно отличается от поведения обычного пользователя. Поэтому для успешной выгрузки необходимо избежать возможной защиты от парсинга, например, блокировки IP-адреса, с помощью имитации парсером поведения обычного пользователя. Например, после выгрузки страницы необходимо установить паузу перед следующим запросом. Важно ставить неравномерные задержки, иначе скрипт, посылающий запросы с частотой, к примеру, ровно 5 секунд, легко отслеживается.

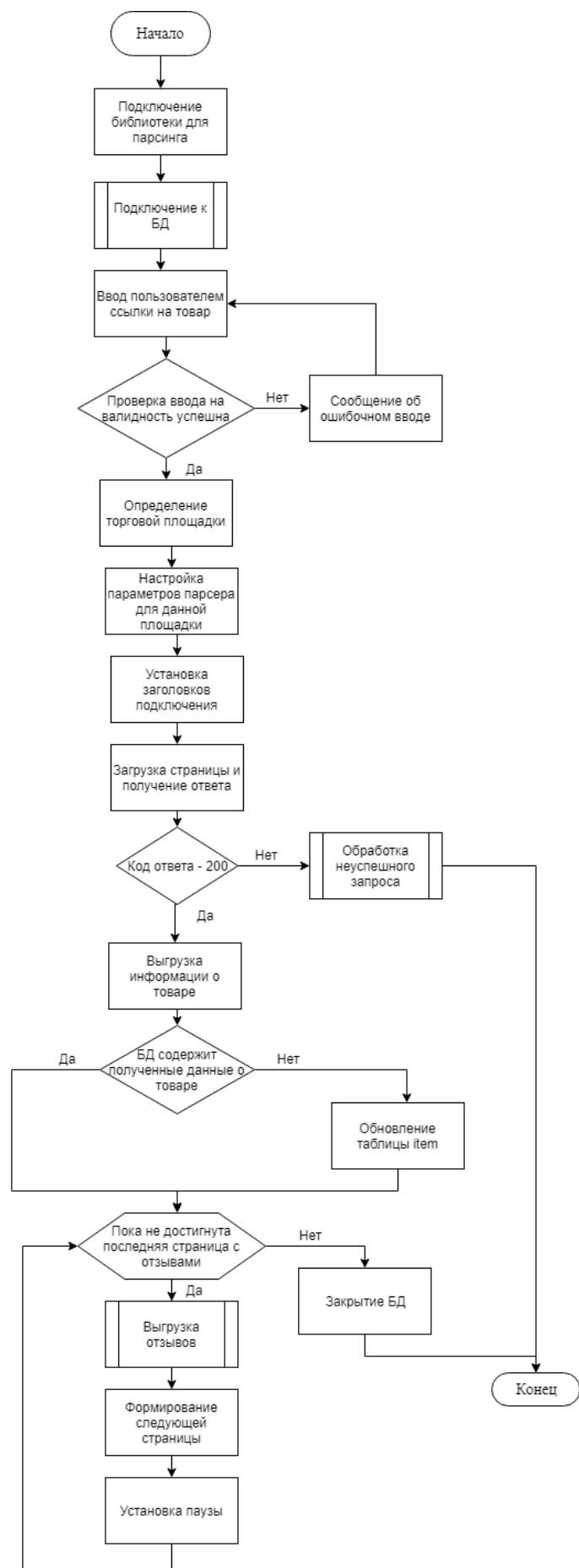


Рисунок 2 – Блок схема главной программы

Выгрузка отзывов также происходит заранее настроенным парсером. В цикле перебираются все отзывы и проверяются на наличие в БД, в случае отсутствия таблица reviews обновляется.

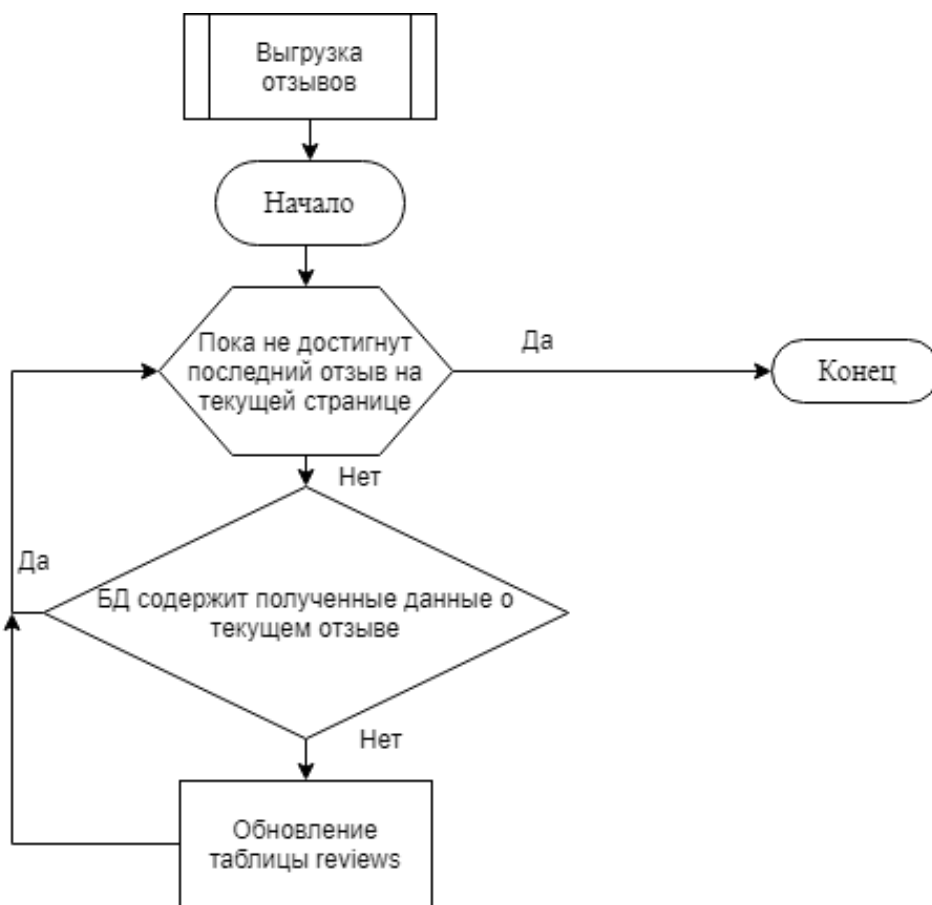


Рисунок 3 – Блок схема функции выгрузки отзывов

## 4 Программная реализация модуля

### 4.1 Язык программирования

В качестве инструмента для разработки был выбран язык PHP. PHP - это распространенный язык программирования общего назначения с открытым исходным кодом. Поскольку наше решение необходимо выполнить кроссплатформенным, модуль будет веб-приложением. Язык PHP специально сконструирован для веб-разработок и его код может внедряться непосредственно в HTML.

PHP является свободным программным обеспечением, распространяемым под особой лицензией. Имеет огромную аудиторию разработчиков, в том числе русское сообщество. PHP имеет отличное взаимодействие с базами данных. PHP предлагает встроенные средства

организации веб-сессий, программный интерфейс расширений. Преимуществами использования данного языка являются широкие возможности и закрытый для пользователей код.

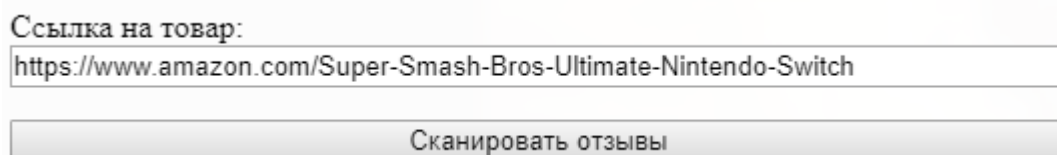
Используемая версия языка – 7.3.5.

## 4.2 Веб-сервер

Для загрузки веб-страницы браузер отправляет запрос на сервер, в пространстве памяти которого хранятся все документы. В текущей работе используется веб-сервер Apache. Apache — свободный веб-сервер с открытым исходным кодом, является кроссплатформенным ПО, поддерживает операционные системы Linux, BSD, Mac OS, Microsoft Windows, Novell NetWare, BeOS. Основными достоинствами Apache считаются надёжность и гибкость конфигурации. Он позволяет подключать внешние модули для предоставления данных, использовать СУБД для аутентификации пользователей, модифицировать сообщения об ошибках и так далее.

## 4.3 HTTP-запросы

После того, как пользователь вставит ссылку на интересующий его товар, модуль отправляет HTTP-запрос для получения данных с указанной площадки.



The image shows a web form with a label 'Ссылка на товар:' followed by a text input field containing the URL 'https://www.amazon.com/Super-Smash-Bros-Ultimate-Nintendo-Switch'. Below the input field is a button labeled 'Сканировать отзывы'.

Рисунок 4 – Ввод ссылки на товар пользователем

Для загрузки страницы с товаром и отзывами необходимо установить заголовки HTTP-запроса. Заголовки — это набор пар имя-значение, разделенных двоеточием. В заголовках передается различная служебная

информация: кодировка сообщения, название и версия браузера, адрес, с которого пришел клиент и так далее.

Основой для создания парсера является библиотека QueryList, с помощью которой и будут устанавливаться необходимые заголовки. Ниже представлены используемые заголовки.

```
1. $headers = [  
2.     'proxy' => $proxies[array_rand($proxies)],  
3.     'timeout' => 180,  
4.     'headers' => [  
5.         'Referer' => $referer,  
6.         'User-Agent' => $user_agent_list[array_rand($user_agent_list)],  
7.         'Accept' => 'application/json',  
8.         'X-Foo' => ['Bar', 'Baz'],  
9.         'Cookie' => 'abc=111;xxx=222'  
10.    ]
```

#### 4.4 Выгрузка необходимых данных

Рассмотрим простой пример для иллюстрации работы библиотеки QueryList при извлечении необходимых данных. Ниже представлен код с извлечением всех имен пользователей, оставивших отзывы на указанной странице торговой площадки «Amazon».

Переменная \$reg хранит параметры для поиска данных, в нашем случае это селектор CSS по классу .a-profile-name и атрибут text, указывающий на необходимость извлечения текста внутри выбранного тега. Помимо текста можно извлекать внутренние фрагменты HTML-кода или любой атрибут выбранного тега.

```
1. $reg = [  
2.     'username' => ['.a-profile-name', 'text'],  
3. ];  
4.  
5. $ql = QueryList::get($page)->rules($reg)->query();  
6.  
7. print_r ($data = $ql->getData());
```

Результат выполнения данного примера представлен на рисунке 5.

```
Tightenco\Collect\Support\Collection Object ( [items:protected] => Array ( [0] => Array ( [username] => Gabe in Oregon )  
[1] => Array ( [username] => 316 ) [2] => Array ( [username] => Amazon Customer ) [3] => Array ( [username] => C Flo  
[4] => Array ( [username] => Kyle Jennings ) [5] => Array ( [username] => Tykhon ) [6] => Array ( [username] => Bryan  
[7] => Array ( [username] => christopher j. ) [8] => Array ( [username] => Richard Munroe ) [9] => Array ( [username]  
=> A.W. Larrison ) [10] => Array ( [username] => Ponch ) [11] => Array ( [username] => Cheryl L. Hoppe ) ) )
```

Рисунок 5 – Результат извлечения имен пользователей

Таким образом, для создания парсера любой торговой площадки необходимо изучить структуру этого веб-ресурса и написать соответствующие параметры поиска, используя необходимые CSS селекторы. Процесс анализа структуры страницы для составления параметров парсинга представлен на рисунке 6.

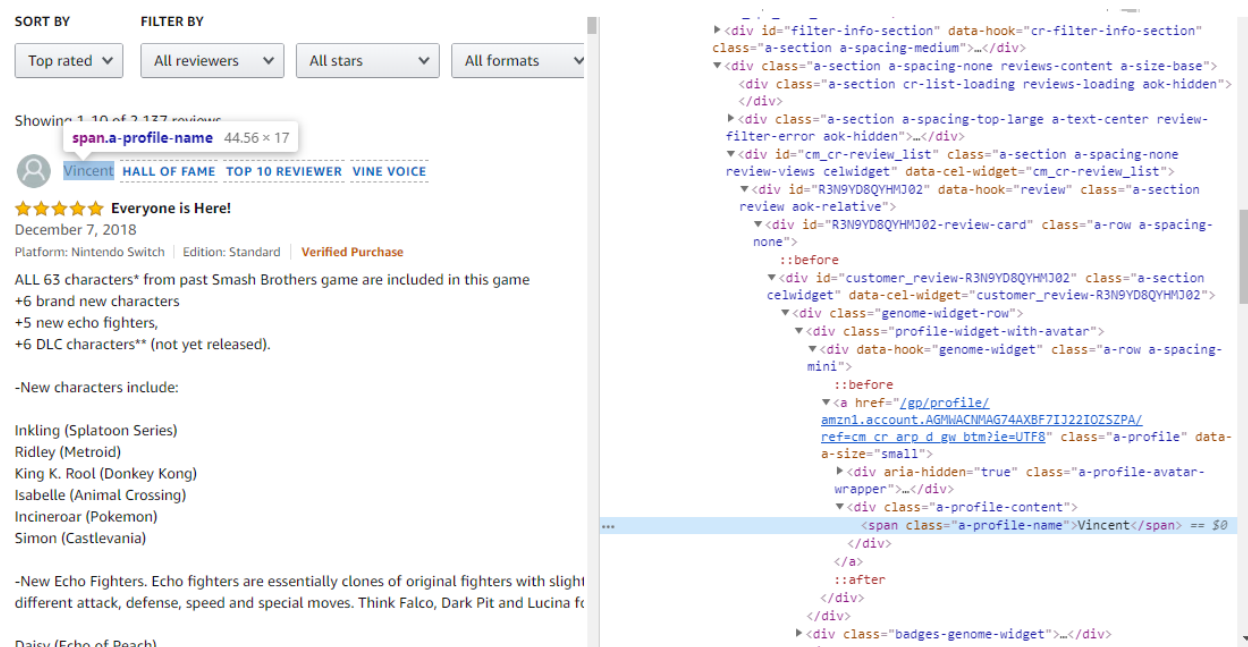


Рисунок 6 – Результат извлечения имен пользователей

## 4.5 Проектирование базы данных

Данные с веб-ресурса загружаются в реляционную базу данных. Проектирование базы данных является также важным этапом, поскольку правильная структура позволит быстрее сохранять и извлекать данные.

При проектировании любой базы данных необходимо соблюдать следующие требования:

- 1) Целостность – полнота и непротиворечивость данных;
- 2) Многократное использование данных;
- 3) Быстрый поиск и получение информации по запросам;
- 4) Простота обновления данных;
- 5) Уменьшение излишней избыточности данных;
- 6) Защита данных от несанкционированного доступа, искажения и уничтожения.



База данных, к которой будет подключаться модуль, создана с помощью возможностей свободной реляционной системы управления базами данных MySQL. Гибкость СУБД MySQL обеспечивается поддержкой большого количества типов таблиц. Как правило, эту систему управления базами данных определяют, как хорошую, быструю и гибкую систему, рекомендованную к применению в небольших или средних проектах.

Структура разработанных таблиц с информацией о товарах и отзывах представлена на рисунках 7 и 8.

#	Имя	Тип	Сравнение	Атрибуты	Null	По умолчанию	Комментарии	Дополнительно
<input type="checkbox"/> 1	<b>id</b>	bigint(20)			Нет	Нем		AUTO_INCREMENT
<input type="checkbox"/> 2	<b>name</b>	tinytext	utf8mb4_0900_ai_ci		Нет	Нем		
<input type="checkbox"/> 3	<b>image_url</b>	varchar(200)	utf8mb4_0900_ai_ci		Нет	Нем		
<input type="checkbox"/> 4	<b>overall_rating</b>	bigint(20)			Нет	Нем		

Рисунок 7 – Таблица item

#	Имя	Тип	Сравнение	Атрибуты	Null	По умолчанию	Комментарии	Дополнительно
<input type="checkbox"/> 1	<b>Item_id</b>	bigint(20)			Нет	Нем		
<input type="checkbox"/> 2	<b>Identifier</b>	bigint(20)			Нет	Нем		AUTO_INCREMENT
<input type="checkbox"/> 3	<b>Rating</b>	int(11)			Нет	Нем		
<input type="checkbox"/> 4	<b>Date</b>	datetime			Нет	Нем		
<input type="checkbox"/> 5	<b>Username</b>	tinytext	utf8mb4_0900_ai_ci		Нет	Нем		
<input type="checkbox"/> 6	<b>Title</b>	varchar(100)	utf8mb4_0900_ai_ci		Нет	Нем		
<input type="checkbox"/> 7	<b>Content</b>	text	utf8mb4_0900_ai_ci		Нет	Нем		
<input type="checkbox"/> 8	<b>Helpful_count</b>	int(11)			Нет	Нем		
<input type="checkbox"/> 9	<b>Verified</b>	tinyint(1)			Нет	Нем		

Рисунок 8 – Таблица reviews

# **ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ**

## **1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения**

### **1.1 Потенциальные потребители результатов исследования**

Данный раздел посвящен актуальности и описанию разрабатываемого модуля извлечения данных с сайтов торговых площадок.

Доля Интернет-торговли в совокупном объеме розничной торговли в мире постепенно растет, в России по данным сайта «РБК» эта доля увеличилась с 3,47% в 2017 году до 4,07% в 2018 году. Согласно результатам прогноза Института экономической политики имени Е.Т. Гайдара, к 2024 году электронная торговля в России может занять 8,5% от всего оборота розничной торговли. Одним из самых важных факторов, влияющих на решение при покупке товара, является его рейтинг среди других покупателей. Именно в рамках Интернет-покупок проще всего обмениваться впечатлениями о приобретенном товаре, любой пользователь может оставить свой отзыв на сайте, давая возможность остальным ознакомиться с ним.

Однако, количество отзывов на товар с некоторых торговых площадок может составлять около 5000 и даже больше. Нынешние объемы информации с отзывами на Интернет-площадках человеку сложно охватить не только для анализа, но и для простого восприятия.

Разрабатываемый сервис предоставляет возможность получения базы данных со структурированной информацией о товарах и отзывах на эти товары с любой электронной торговой площадки. Сейчас существует возможность использования хорошо известных методов математической статистики и машинного обучения для автоматического анализа большого

количества данных. Таким образом, прямыми потребителями данного сервиса являются разработчики систем интеллектуального анализа данных для создания готовой целевой системы. Такая система в результате анализа будет выдавать пользователю рекомендации по покупке интересующего его товара.

## 1.2 Анализ конкурентных технических решений

Детальный анализ конкурирующих разработок, существующих на рынке, необходимо проводить систематически, поскольку рынки пребывают в постоянном движении. Такой анализ помогает вносить коррективы в научное исследование, чтобы успешнее противостоять своим соперникам. Важно реалистично оценить сильные и слабые стороны разработок конкурентов.

В данном пункте рассмотрим 2 конкурентных решения – программу для парсинга различных сайтов Datacol и сервис сбора и агрегации отзывов Mneniya.pro, способных выполнить поставленную задачу, и предлагаемое в данной работе решение. Для проведения сравнения выделим наиболее важные факторы, на которые будут обращать внимание потенциальные пользователи. Результаты анализа представлены в таблице 2.

Таблица 2 – Оценочная карта для сравнения конкурентных технических решений (разработок)

№ п/п	Конкуренты	Факторы конкурентоспособности					Итоговая оценка
		Автоматическая настройка парсера	Удобный экспорт данных в базу данных	Низкая стоимость	База площадок для парсинга	Обработка данных перед сохранением	
1	Datacol	0/0	7/2,31	7/1,4	8/1,04	10/0,7	5,45
2	Mneniya.Pro	10/2,7	3/0,99	3/0,6	5/0,65	5/0,35	5,29
4	Разрабатываемый модуль	10/2,7	10/3,3	10/2	3/0,39	0/0	8,39
	bj [1-5]	4	5	3	2	1	15
	wj [0-1]	0,27	0,33	0,2	0,13	0,07	1



Рисунок 9 – Многоугольник конкурентоспособности

Приведенный анализ конкурентоспособности показал, что разрабатываемый модуль обладает высокой конкурентоспособностью, поскольку имеет высокие показатели по наиболее важным для решения поставленной задачи факторам. Низкие результаты существующих решений объясняются тем, что данные проекты хоть и применимы в рамках нашей задачи, но были созданы для других целей. Так, платформа Mneniya.pro имеет мощный функционал для сбора и публикации отзывов на товары на сайтах, а программа Datacol является универсальным парсером с возможностью настройки парсинга и обработки полученных результатов, который удобнее использовать в качестве отдельного продукта. Разрабатываемый модуль ориентирован именно на дальнейшее включение в систему интеллектуального анализа данных, поэтому автоматически настраивает парсинг и экспортирует все результаты в базу данных. Главным

направлением в сторону улучшения является расширение базы площадок для парсинга, идеальным вариантом для разрабатываемого модуля будет создание универсального парсера.

### 1.3 SWOT-анализ

Данный пункт посвящен созданию SWOT-анализа для разрабатываемого модуля. SWOT-анализ представляет собой комплексный анализ научно-исследовательского проекта и применяется для исследования внешней и внутренней среды проекта. Полученная SWOT-матрица представлена в таблице 3.

Таблица 3 – SWOT-анализ проекта

	Сильные стороны проекта: 1. Расширяемая база торговых площадок для парсинга 2. Удобство использования в других проектах	Слабые стороны проекта: 1. Сложность создания универсального парсера 2. Недостаток выработки требований и проектирования
Возможности: 1. Развитие Интернет-торговли 2. Рост систем интеллектуального анализа данных 3. Большое количество библиотек для парсинга	[2-1]Рост систем интеллектуального анализа данных повысит спрос на базы данных с различных площадок для тестирования систем в процессе разработки и использования уже в реализованной системе.	[3-1] Новые библиотеки предоставят новые возможности для создания универсального парсера
Угрозы: 1. Недостовверные отзывы 2. Усиленная защита от парсинга 3. Конкуренция 4. Частые изменения сайтов торговых площадок	[3-2]Ориентированность на дальнейшее использование модуля в других проектах является преимуществом перед конкурентами.	[3,2,4-2] Аналогичные решения с детальным проектированием будут более устойчивы как к изменению сайтов, так и к защите. Поэтому для повышения конкурентоспособности необходимо детально спроектировать проект.

## 2 Планирование научно-исследовательских работ

### 2.1 Структура работ в рамках научного исследования

Одной из составляющих успешной реализации проекта служит рациональное планирование занятости каждого из его участников, а также определение сроков выполнения определенных этапов работы над проектом. В данном разделе приводится перечень этапов работы, исполнителей, а также оценивается степень участия каждого из участников в том или ином этапе. Данные по перечню работ и продолжительности работ представлены в таблице 4. Число исполнителей данного проекта равно двум – непосредственный исполнитель и научный руководитель.

Таблица 4 – Перечень работ и распределение исполнителей

№ работы	Наименование работы	Исполнители работы
1	Выбор научного руководителя бакалаврской работы	Сутягина А.А.
2	Составление и утверждение темы бакалаврской работы	Саврасов Ф.В., Сутягина А.А.
3	Составление календарного плана-графика выполнения бакалаврской работы	Саврасов Ф.В.
4	Подбор и изучение литературы по теме бакалаврской работы	Сутягина А.А.
5	Выработка требований к разрабатываемому модулю	Саврасов Ф.В., Сутягина А.А.
6	Проектирование архитектуры модуля	Сутягина А.А.
7	Выбор языка разработки и используемых библиотек	Сутягина А.А.
8	Конструирование модуля	Сутягина А.А.
8	Тестирование работы модуля	Сутягина А.А.
9	Согласование выполненной работы с научным руководителем	Саврасов Ф.В., Сутягина А.А.
10	Выполнение других частей работы (финансовый менеджмент, социальная ответственность)	Сутягина А.А.
11	Подведение итогов, оформление работы	Сутягина А.А.



## 2.2 Определение трудоемкости выполнения работ

Трудовые затраты в большинстве случаев образуют основную часть стоимости разработки, поэтому важным моментом является определение трудоемкости работ каждого из участников научного исследования.

Трудоемкость выполнения научного исследования оценивается экспертным путем в человеко-днях и носит вероятностный характер, т.к. зависит от множества трудно учитываемых факторов. Для определения ожидаемого (среднего) значения трудоемкости  $t_{\text{ож}i}$  используется следующая формула:

$$t_{\text{ож}i} = \frac{3t_{\text{min}i} + 2t_{\text{max}i}}{5}, \quad (1)$$

где  $t_{\text{ож}i}$  – ожидаемая трудоемкость выполнения  $i$ -ой работы чел.-дн.;

$t_{\text{min}i}$  – минимально возможная трудоемкость выполнения заданной  $i$ -ой работы (оптимистическая оценка: в предположении наиболее благоприятного стечения обстоятельств), чел.-дн.;

$t_{\text{max}i}$  – максимально возможная трудоемкость выполнения заданной  $i$ -ой работы (пессимистическая оценка: в предположении наиболее неблагоприятного стечения обстоятельств), чел.-дн.

Исходя из ожидаемой трудоемкости работ, определяется продолжительность каждой работы в рабочих днях  $T_p$ , учитывающая параллельность выполнения работ несколькими исполнителями. Такое вычисление необходимо для обоснованного расчета заработной платы, так как удельный вес зарплаты в общей сметной стоимости научных исследований составляет около 65 %.

$$T_{pi} = \frac{t_{\text{ож}i}}{Ч_i}, \quad (2)$$

где  $T_{pi}$  – продолжительность одной работы, раб. дн.;

$t_{\text{ож}i}$  – ожидаемая трудоемкость выполнения одной работы, чел.-дн.;

$Ч_i$  – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел.

### 2.3 Разработка графика проведения научного исследования

Для удобства построения графика, длительность каждого из этапов работ из рабочих дней следует перевести в календарные дни. Для этого необходимо воспользоваться следующей формулой:

$$T_{ki} = T_{pi} \cdot k_{\text{кал}}, \quad (3)$$

где  $T_{ki}$  – продолжительность выполнения  $i$ -й работы в календарных днях;

$T_{pi}$  – продолжительность выполнения  $i$ -й работы в рабочих днях;

$k_{\text{кал}}$  – коэффициент календарности.

Коэффициент календарности определяется по следующей формуле:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - T_{\text{вых}} - T_{\text{пр}}}, \quad (4)$$

где  $T_{\text{кал}}$  – количество календарных дней в году;

$T_{\text{вых}}$  – количество выходных дней в году;

$T_{\text{пр}}$  – количество праздничных дней в году.

Согласно производственному календарю (для 6-дневной рабочей недели) в 2019 году 365 календарных дней, 299 рабочих дней, 66 выходных/праздничных дней.

$$K_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - T_{\text{вых}} - T_{\text{пр}}} = \frac{365}{365 - 66} = 1,22$$

С помощью указанных в пункте 2.2 расчетных формул вычислим для каждой работы ее трудоемкость и длительность. По полученных результатам построим график Ганта.

Таблица 5 – Временные показатели проведения научного исследования

№	Наименование работы	Исполнители работы	Трудоемкость работ, чел-дни			Длительность работ, дни	
			tmin	tmax	тож	Тр	Тк
1	Выбор научного руководителя бакалаврской работы	Сутягина А.А.	1	3	1.8	2	2
2	Составление и утверждение темы бакалаврской работы	Сутягина А.А.	1	2	1.4	1	1
		Саврасов Ф.В.	1	2	1.4	1	1
3	Составление календарного плана-графика выполнения бакалаврской работы	Саврасов Ф.В.	1	1	1	1	1
4	Подбор и изучение литературы по теме бакалаврской работы	Сутягина А.А.	3	5	3.8	4	5
		Саврасов Ф.В.	1	1	1	1	1
5	Выработка требований к разрабатываемому модулю	Саврасов Ф.В.	1	1	1	1	1
		Сутягина А.А.	2	5	3.2	3	4
6	Проектирование архитектуры модуля	Сутягина А.А.	5	10	7	7	9
		Саврасов Ф.В.	1	1	1	1	1
7	Выбор языка разработки и используемых библиотек	Сутягина А.А.	3	5	3.8	4	5
		Саврасов Ф.В.	1	2	1.4	1	1
8	Конструирование модуля	Сутягина А.А.	40	70	52	52	63
		Саврасов Ф.В.	1	4	2.2	2	2
9	Тестирование работы модуля	Сутягина А.А.	3	7	4.6	5	6
		Саврасов Ф.В.	1	2	1.4	1	1
10	Согласование выполненной работы с научным руководителем	Сутягина А.А.	1	5	2.6	3	4
		Саврасов Ф.В.	1	3	1.8	2	2
11	Выполнение других частей работы (финансовый менеджмент, социальная ответственность)	Сутягина А.А.	8	13	10	10	12
12	Подведение итогов, оформление работы	Сутягина А.А.	3	7	4.6	5	6

Таблица 6 – График Ганта

	01.02	03.02	04.02	05.02	06.02	10.02	11.02	14.02	15.02	23.02	24.02	28.02	02.03	02.05	03.05	08.05	10.05	12.05	24.05	30.05
1	АА																			
2		АА																		
3		ФВ																		
4				АА																
5				ФВ		АА														
6								АА												
7								ФВ												
8										АА										
9														АА						
10														ФВ		АА				
11																		АА		
12																			АА	
АА													Сутягина А.А.							
ФВ													Саврасов Ф.В.							

Таким образом, общая длительность работы над проектом составляет 117 дней.

## **2.4 Бюджет научно-технического исследования**

Произведем расчет необходимого для проведения всех работ бюджета.

### **2.4.1 Расчет материальных затрат научно-технического исследования**

Данная статья включает в себя расчет затрат на все материалы, используемые при разработке проекта.

В материальные затраты включаются затраты на канцелярские принадлежности, диски, картриджи и т.п. Материальные затраты составляют канцелярские расходы в сумме 500 рублей.

### **2.4.2 Расчет затрат на специальное оборудование для научных (экспериментальных) целей**

Данная статья включает в себя расчет затрат на приобретение необходимого технического оборудования.

Произведем расчет амортизации используемого персонального компьютера. Первоначальная стоимость ПК составляет 40 000 рублей, срок полезного использования для машин офисных код 330.28.23.23 составляет 2-3 года, берем 3 года, планируем использовать ПК для написания ВКР в течение 4 месяцев. Тогда:

- норма амортизации:

$$A_n = \frac{1}{n} * 100\% = \frac{1}{3} \times 100\% = 33,33\%$$

- годовые амортизационные отчисления:

$$A_g = 40000 \times 0,33 = 13\,200 \text{ рублей}$$

- ежемесячные амортизационные отчисления:

$$A_m = \frac{13200}{12} = 1\,100 \text{ рублей}$$

- итоговая сумма амортизации основных средств:

$$A = 1100 \times 4 = 4\,400 \text{ рублей}$$

Значит, в затраты на специальное оборудование необходимо включить сумму амортизации основных средств в сумме 4 400 руб.

### 2.4.3 Основная заработная плата исполнителей темы

Данная статья включает в себя расчет заработной платы всех исполнителей темы. Заработная плата рассчитывается исходя из трудоемкости и действующей системы окладов в ТПУ. В состав основной заработной платы также включаются премии и доплаты.

Для начала необходимо определить действительный годовой фонд рабочего времени.

Таблица 7 – Баланс рабочего времени (для 6-дневной недели)

Показатели рабочего времени	Дни
Календарные дни	365
Нерабочие дни (праздники/выходные)	66
Потери рабочего времени (отпуск/невыходы по болезни)	56
Действительный годовой фонд рабочего времени	243

Для расчета основной заработной платы студента берем оклад, равный окладу ассистента без степени, т.е. 21 760 руб.

Для расчета основной заработной платы руководителя берем оклад в соответствии с занимаемой должностью и наличием ученой степени. Оклад доцента кандидата наук составляет 33 664 руб.

Произведем вычисление затрат на заработную плату:

$$Зп = Зосн + Здоп, \quad (5)$$

где  $Z_{осн}$  – основная заработная плата, руб.;

$Z_{доп}$  – дополнительная заработная плата, руб.

Вычислим среднедневную заработную плату:

$$Z_{дн} = \frac{Z_{м \times M}}{F_{д}}, \quad (6)$$

где  $Z_{м}$  – месячный оклад работника, руб.

$M$  – количество месяцев работы без отпуска в течение года:

при отпуске в 24 раб. дня  $M = 11,2$  месяца, 5-дневная неделя;

при отпуске в 48 раб. дней  $M = 10,4$  месяца, 6-дневная неделя;

$F_{д}$  – действительный годовой фонд рабочего времени персонала, раб.

дни.

Среднедневная заработная плата научного руководителя:

$$Z_{дн} = \frac{Z_{м} \times M}{F_{д}} = \frac{33\,664 \times 10,4}{243} = 1\,440,76 \text{ руб.}$$

Среднедневная заработная плата студента:

$$Z_{дн} = \frac{Z_{м} \times M}{F_{д}} = \frac{21\,760 \times 10,4}{243} = 931,29 \text{ руб.}$$

Вычислим основную заработную плату:

$$Z_{осн} = Z_{дн} \times T_{р} \times (1 + K_{пр} + K_{д}) \times K_{р} \quad (7)$$

где  $Z_{дн}$  – среднедневная заработная плата, руб.;

$K_{пр}$  – премиальный коэффициент;

$K_{д}$  – коэффициент доплат и надбавок;

$K_{р}$  – районный коэффициент;

$T_{р}$  – продолжительность работ, выполняемых работником, раб. дни.

Основная заработная плата научного руководителя:

$$Z_{осн} = 1\,440,76 \times 11 \times (1 + 0,3 + 0,35) \times 1,3 = 33\,994,73 \text{ руб.}$$

Основная заработная плата студента:

$$Z_{осн} = 931,29 \times 96 \times (1 + 0,3 + 0,35) \times 1,3 = 191\,771,24 \text{ руб.}$$

Таблица 8 – Расчет основной заработной платы

Исполнители	Здн, руб.	Кпр	Кд	Кр	Тр	Зосн
Научный руководитель	1 440,76	0,3	0,35	1,3	11	33 994,73
Студент	931,29	0,3	0,35	1,3	96	191 771,24
Итого:						225 765,97

#### 2.4.4 Дополнительная заработная плата исполнителей темы

Затраты по дополнительной заработной плате исполнителей темы учитывают величину предусмотренных Трудовым кодексом РФ доплат за 74 отклонение от нормальных условий труда, а также выплат, связанных с обеспечением гарантий и компенсаций (при исполнении государственных и общественных обязанностей, при совмещении работы с обучением, при предоставлении ежегодного оплачиваемого отпуска и т.д.).

Расчет дополнительной заработной платы ведется по следующей формуле:

$$З_{\text{доп}} = k_{\text{доп}} \cdot З_{\text{осн}} \quad (8)$$

где  $k_{\text{доп}}$  – коэффициент дополнительной заработной платы (на стадии проектирования принимается равным 0,12 – 0,15).

В данной работе примем его равным 0,12.

Для научного руководителя:

$$З_{\text{доп}_\Pi} = 33\,994,73 \cdot 0,12 = 4\,079,37 \text{ рублей}$$

Для студента:

$$З_{\text{доп}_\text{С}} = 191\,771,24 \cdot 0,12 = 23\,012,55 \text{ рублей}$$

Дополнительная заработная плата научного руководителя: 4 079,37 руб.;

Дополнительная заработная плата студента: 23 012,55 руб.;

Итого: 27 091,92 руб.



### **2.4.5 Отчисления во внебюджетные фонды (страховые отчисления)**

В данной статье расходов отражаются обязательные отчисления по установленным законодательством Российской Федерации нормам органам государственного социального страхования (ФСС), пенсионного фонда (ПФ) и медицинского страхования (ФФОМС) от затрат на оплату труда работников. Размер страховых взносов составляет 30% от основной и дополнительной заработных плат.

Величина отчислений во внебюджетные фонды определяется исходя из следующей формулы:

$$З_{\text{внеб}} = k_{\text{внеб}} \cdot (З_{\text{осн}} + З_{\text{доп}}) \quad (9)$$

где  $k_{\text{внеб}}$  – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.).

Для научного руководителя:

$$З_{\text{внеб\_П}} = (33\,994,73 + 4\,079,37) \cdot 0,3 = 11\,422,23 \text{ рублей}$$

Для студента:

$$З_{\text{внеб\_С}} = (191\,771,24 + 23\,012,55) \cdot 0,3 = 64\,435,14 \text{ рублей}$$

Страховые отчисления научного руководителя: 11 422,23 руб.;

Страховые отчисления студента: 64 435,14 руб.;

Итого: 75 857,37 руб.

### **2.4.6 Накладные расходы**

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: печать и ксерокопирование материалов исследования, оплата услуг связи, электроэнергии, почтовые и телеграфные расходы, размножение материалов и т.д. Их величина определяется по следующей формуле:

$$\text{Знакл} = (\text{сумма расходов статей}) \cdot \text{кнр}, \quad (10)$$

где кнр – коэффициент, учитывающий накладные расходы.

Величину коэффициента накладных расходов можно взять в размере 16%.

$$\begin{aligned} \text{З}_{\text{накл}} &= (500 + 4\,400 + 225\,765,97 + 27\,091,92 + 75\,857,37) \cdot 0,16 \\ &= 53\,378,44 \text{ рублей} \end{aligned}$$

Накладные расходы составили 53 378,44 руб.

#### 2.4.7 Формирование бюджета затрат научно-исследовательского проекта

Используя полученные в предыдущих пунктах результаты, получим общий бюджет и рассчитаем удельный вес в процентах каждой составляющей.

Таблица 9 – Расчет бюджета затрат НТИ

Наименование	Сумма, руб.	Удельный вес, %
Материальные затраты	500	0,13%
Затраты на специальное оборудование	4 400	1,14%
Затраты на основную заработную плату	225 765,97	58,34%
Затраты на дополнительную заработную плату	27 091,92	7,00%
Страховые взносы	75 857,37	19,60%
Накладные расходы	53 378,44	13,79%
Общий бюджет	386 993.7	100%

### 3 Определение потенциального эффекта исследования

В результате проведенной работы были определены конечные потребители разрабатываемого модуля, изучены их потребности, выявлены значимые факторы, на основе которых проведен анализ конкурентных решений. Данный анализ указал на необходимость создания универсального модуля, способного работать с любым сайтом в любое время. Последующий SWOT-анализ дал решение этой проблемы в изучении новых библиотек для парсинга, а также указал на необходимость детального проектирования разработки.

Проведенное планирование позволяет определить конкретные сроки выполнения всех необходимых этапов работы для того, чтобы вся работа была выполнена к нужному дню. Дальнейшее следование полученному плану позволит избежать непредвиденных задержек.

Формирование бюджета затрат сделает возможным предварительную подготовку всех необходимых средств и позволит избежать незапланированных расходов.

По результатам проведенной оценки данная разработка признана конкурентоспособной и перспективной. Общая длительность разработки составила 117 календарного дня. Планируемая стоимость данной разработки составила около 387 тыс. рублей.

# **СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ**

## **Введение**

Данный раздел посвящен вопросам обеспечения безопасности при проектировании объекта исследования и при его последующей эксплуатации. В ходе данной работы разрабатывается Web-модуль для парсинга данных с сайтов торговых площадок, поэтому необходимо рассмотреть все аспекты работы с ПЭВМ, как со стороны влияния на здоровье пользователей, так и со стороны влияния на окружающую среду. Прямыми пользователями являются разработчики систем интеллектуального анализа данных, использующие базу данных, создаваемую разрабатываемым модулем.

## **1 Правовые и организационные вопросы обеспечения безопасности.**

### **1.1 Специальные (характерные для проектируемой рабочей зоны) правовые нормы трудового законодательства.**

В СанПиН 2.2.2/2.4.1340-03 содержатся следующие правила и нормы, описывающие проектируемую рабочую зону:

- Продолжительность рабочего дня не должна превышать 40 часов в неделю;
- Продолжительность непрерывной работы за компьютером без регламентированного перерыва не должна превышать 1 час;
- Рекомендуется делать перерывы в работе за ПК продолжительностью 10-15 минут через каждые 45-60 минут работы;
- Во время регламентированных перерывов целесообразно выполнять комплексы упражнений и осуществлять проветривание помещения;
- Не рекомендуется работать за компьютером более 6 часов за смену.

Существуют также специализированные органы, осуществляющие государственный контроль и надзор в организациях на предмет соблюдения существующих правил и норм. К таким органам относятся:

- Федеральная инспекция труда;
- Государственная экспертиза условий труда Федеральной службы по труду и занятости населения;
- Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека и др.

## **1.2 Организационные мероприятия при компоновке рабочей зоны.**

Большое значение для профилактики статических физических перегрузок имеет правильная организация рабочего места человека, работающего с ПЭВМ. Рабочее место должно быть организовано в соответствии с требованиями стандартов, технических условий и (или) методических указаний по безопасности труда. Оно должно удовлетворять следующим требованиям:

- обеспечивать возможность удобного выполнения работ;
- учитывать физическую тяжесть работ;
- учитывать размеры рабочей зоны и необходимость передвижения в ней работающего;
- учитывать технологические особенности процесса выполнения работ.

Невыполнение требований к расположению и компоновке рабочего места может привести к получению пользователем производственной травмы или развития у него профессионального заболевания. Рабочее место программиста должно соответствовать требованиям СанПин 2.2.2/2.4.1340-03.

Конструкция оборудования и рабочего места при выполнении работ в положении сидя должна обеспечивать оптимальное положение работающего, которое достигается регулированием высоты рабочей поверхности, высоты сидения, оборудованием пространства для размещения ног и высотой подставки для ног. Схемы размещения рабочих мест с персональными компьютерами должны учитывать расстояния между рабочими столами с мониторами: расстояние между боковыми поверхностями мониторов не менее 1,2 м, а расстояние между экраном монитора и тыльной частью другого монитора не менее 2,0 м. Клавиатура должна располагаться на поверхности стола на расстоянии 100-300 мм от края, обращенного к пользователю. Быстрое и точное считывание информации обеспечивается при расположении плоскости экрана ниже уровня глаз пользователя, предпочтительно перпендикулярно к нормальной линии взгляда (нормальная линия взгляда 15 градусов вниз от горизонтали). Рабочие места с ПЭВМ при выполнении творческой работы, требующей значительного умственного напряжения или высокой концентрации внимания, рекомендуется изолировать друг от друга перегородками высотой 1,5 - 2,0 м.

Соблюдение указанных норм позволит избежать проблем со здоровьем, что в свою очередь способствует повышению работоспособности пользователей ПЭВМ.

## **2 Профессиональная социальная безопасность.**

### **2.1 Анализ вредных и опасных факторов при проведении исследований и при использовании объекта исследования и обоснование мероприятий по защите исследователя.**

Поскольку рабочее место при проведении исследований и при использовании объекта исследования одно и то же, помещение с ПЭВМ,

описание вредных и опасных факторов для обоих случаев поместим в один пункт.

Вредным называется производственный фактор, воздействие которого на работающего в определенных условиях приводит к заболеванию или снижению работоспособности. При изменении уровня и времени воздействия вредные производственные факторы могут стать опасными. Опасными считаются производственные факторы, воздействие которых на работающего в конкретных условиях может привести к травмам, а также другим внезапным резким ухудшениям здоровья.

При работе с ПЭВМ пользователь подвергается воздействию опасных и вредных производственных факторов. Основные из них:

1. электромагнитные поля;

Источниками электромагнитного излучения в данном исследовании являются мониторы и системный блок.

Электромагнитное излучение может вызвать расстройства нервной системы, снижение иммунитета, расстройства сердечно-сосудистой системы и представляет собой особую опасность для беременных.

Предельно-допустимые нормы электромагнитного излучения при работе на компьютере представлены в таблице 910, источником является СанПиН 2.2.2/2.4.1340-03.

Таблица 10 – Предельно-допустимые нормы ЭМП

Напряженность электрического поля	
В диапазоне частот 5 Гц — 2 кГц	25 В/м
В диапазоне частот 2 кГц — 400 кГц	2,5 В/м
Плотность магнитного потока	
В диапазоне частот 5 Гц — 2 кГц	250 нТл
В диапазоне частот 2 кГц — 400 кГц	25 нТл

Для того, чтобы снизить воздействие таких видов излучения, рекомендуют применять такие мониторы, у которых уровень излучения понижен (MPR-II, TCO-92, TCO-99), а также установить защитные экраны и соблюдать режимы труда и отдыха.

## 2. электростатические поля;

При включенном питании компьютера на экране дисплея накапливается статическое электричество.

Электрический ток искрового разряда статического электричества мал и не может вызвать поражение человека. Тем не менее, вблизи экрана электризуется пыль и оседает на нем. В результате чего искажается резкость восприятия информации на экране. Кроме того, пыль попадает на лицо работающего и в его дыхательные пути.

Согласно ГОСТ 12.1.045–84 предельно допустимый уровень напряженности электростатических полей устанавливается равным 60 кВ/м в течение 1 часа.

Основные способы защиты от статического электричества следующие: заземление оборудования, увлажнение окружающего воздуха. Также целесообразно применение полов из антистатического материала.

## 3. шум;

Работа ПЭВМ сопровождается акустическими шумами. Источниками шума ПЭВМ являются:

- Кулер на процессоре;
- Кулеры на корпусе;
- Система охлаждения видеокарты;
- Жесткие диски;
- CD-ROM.

Техногенные шумы оказывают негативные последствия для здоровья и психики человека. Примером такого отрицательного влияния является потеря общего тонуса, сонливость и головные боли.

В соответствии с требованиями санитарных норм СН 2.2.4/2.1.8.562-96 предельно допустимые уровни звука для разных категорий напряженности трудового процесса легкой и средней физической нагрузки представлены в таблице 11.



Таблица 11 – Предельно допустимые уровни звука и эквивалентные уровни звука на рабочих местах для трудовой деятельности разных категорий напряженности

Категория напряженности трудового процесса	Предельно допустимый уровень звука, дБА
Напряженность легкой степени	80
Напряженность средней степени	70
Напряженный труд 1 степени	60
Напряженный труд 2 степени	50

Защита от шумов – заключение вентиляторов в защитный кожух и установление их внутри корпуса ПЭВМ. Для снижения уровня шума стены и потолок помещений, где установлены компьютеры, могут быть облицованы звукопоглощающими материалами с максимальными коэффициентами звукопоглощения в области частот 63 - 8000 Гц.

#### 4. нарушения микроклимата в помещении;

Микроклимат рабочих помещений – метеорологические условия внутренней среды помещений, которые определяются действующими на организм человека сочетаниями температуры, влажности, скорости движения воздуха и теплового излучения; комплекс физических факторов, оказывающих влияние на теплообмен человека с окружающей средой, на тепловое состояние человека и определяющих самочувствие, работоспособность, здоровье и производительность труда.

Нормы оптимальных и допустимых показателей микроклимата при работе с ЭВМ устанавливает СанПиН 2.2.4.548-96 «Гигиенические требования к микроклимату производственных помещений» и ССБТ ГОСТ 12.1.005-88 «Общие санитарно-гигиенические требования к воздуху рабочей зоны». Все категории работ разграничиваются на основе интенсивности энергозатрат организма в ккал/ч (Вт). Работа, производимая сидя и сопровождающаяся незначительным физическим напряжением, относится к

категории Ia – работа с интенсивностью энергозатрат до 120 ккал/ч (до 139 Вт).

Допустимый уровень микроклимата помещения обеспечивается системой водяного центрального отопления и естественной вентиляцией. В помещениях, где допустимые нормативные величины микроклимата поддерживать не представляется возможным, необходимо проводить мероприятия по защите работников от возможного перегрева и охлаждения. Это достигается различными средствами, например, применением систем местного кондиционирования воздуха, использованием индивидуальных средств защиты от повышенной или пониженной температуры, регламентацией периодов работы в неблагоприятном микроклимате и отдыха в помещении с микроклиматом, нормализующим тепловое состояние, сокращением рабочей смены и др.

Таблица 12 – Оптимальные нормы микроклимата

Период года	Категория работ по уровню энергозатрат, Вт	Температура, °С	Относительная влажность, %	Скорость движения воздуха, м/с
Холодный	Ia	(22÷24)	(40÷60)	0,1
Теплый	Ia	(23÷25)	(40÷60)	0,1

Таблица 13 – Допустимые нормы микроклимата

Период года	Категория работ по уровню энергозатрат, Вт	Температура, °С		Относительная влажность, %	Скорость движения воздуха, м/с	
		Диапазон ниже оптимальных величин	Диапазон выше оптимальных величин		Для диапазона температур ниже оптимальных величин, не более	Для диапазона температур выше оптимальных величин, не более
Холодный	Ia	(20,0÷21,9)	(24,1÷25,0)	(15÷75)	0,1	0,1
Теплый	Ia	(21,0÷22,9)	(25,1÷28,0)	(15÷75)	0,1	0,2

#### 5. нарушения освещенности рабочей зоны;

Освещение – получение, распределение и использование световой энергии для обеспечения благоприятных условий видения предметов и объектов. Оно влияет на настроение и самочувствие, определяет эффективность труда.

Осветительные установки, независимо от используемых источников света и световых приборов, должны обеспечивать нормативные требования к общему искусственному освещению. В помещениях с ПЭВМ нормируемый показатель искусственного освещения согласно СанПиН 2.2.1/2.1.1.1278–03 равен 300 лк.

#### 6. психофизиологические факторы.

Психофизиологические факторы вредного воздействия включают в себя стресс, переутомление, монотонность труда. Нерациональное расположение элементов рабочего места вызывает необходимость поддержания вынужденной рабочей позы. Длительный дискомфорт вызывает повышенное позвоночное напряжение мышц и обуславливает развитие общего утомления и снижение работоспособности.

При длительной работе за экраном дисплея появляется выраженное напряжение зрительного аппарата с появлением жалоб на неудовлетворительность работы, головные боли, усталость и болезненное ощущение в глазах, в пояснице, в области шеи, руках.

Режим труда и отдыха работника: при вводе данных, редактировании программ, чтении информации с экрана непрерывная продолжительность работы не должна превышать 4-х часов при 8-часовом рабочем дне. Через каждый час работы необходимо делать перерыв на 5-10 минут, а через два часа – на 15 минут.

С целью снижения или устранения нервно-психологического, зрительного и мышечного напряжения, предупреждение переутомления необходимо проводить комплекс физических упражнений и сеансы

психофизической разгрузки и снятия усталости во время регламентируемых перерывов, и после окончания рабочего дня.

Воздействие перечисленных факторов уменьшает работоспособность человека, а длительное и систематическое воздействие может приводить к профессиональному заболеванию.

### **3 Экологическая безопасность.**

#### **3.1 Анализ влияния процесса исследования и объекта исследования на окружающую среду.**

Современное общество сопровождается бурным развитием промышленности, что влечёт за собой необратимые последствия в виде загрязнений окружающей среды.

Влияние на окружающую среду при проведении исследования и при дальнейшем использовании объекта исследования проявляется в:

- увеличении энергопотребления;
- отсутствии утилизации ПЭВМ;
- отсутствие утилизации макулатуры.

Все это приводит к избыточному вмешательству человека в процессы окружающей среды, разрушая ее и не давая возможности восстановиться.

#### **3.2 Обоснование мероприятий по защите окружающей среды.**

Так как основную часть работы занимает разработка с использованием ПЭВМ, серьезной проблемой является электропотребление. Для удовлетворения потребности в электроэнергии, приходится увеличивать мощность и количество электростанций. Это приводит к нарушению экологической обстановки, так как электростанции в своей деятельности используют различные виды топлива, водные ресурсы, а также

являются источником вредных выбросов в атмосферу. Однако использование разрабатываемого модуля позволит заменить многократные обращения к Интернет-площадкам для выгрузки отзывов с разных страниц одним обращением к базе данных.

При рассмотрении влияния процесса утилизации ПЭВМ были выявлены особо вредные выбросы согласно ГОСТ Р 51768-2001. В случае выхода из строя компьютеров, они списываются и отправляются на специальный склад, который при необходимости принимает меры по утилизации списанной техники и комплектующих. В настоящее время в Томской области утилизацией занимаются две компании: городской полигон и ООО НПП «Экотом». Утилизацией опасных бытовых отходов занимаются компании: ООО «Торем», ООО «СибМеталлГрупп».

Основной вид мусора – это отходы печати, коробки от техники, использованная бумага. В Томске есть большое количество пунктов приема макулатуры.

#### **4 Безопасность в чрезвычайных ситуациях.**

##### **4.1 Анализ вероятных ЧС, которые могут возникнуть на рабочем месте при проведении исследований или дальнейшем использовании объекта исследований.**

Наиболее вероятной ЧС в рассматриваемом случае является пожар (взрыв) в здании. Источниками возгорания может стать электропроводка, внутренние работающие устройства ПЭВМ, взрывоопасные предметы в помещении исследователя согласно ГОСТ 12.1.044-89 «Система стандартов безопасности труда. Пожаровзрывоопасность веществ и материалов. Номенклатура показателей и методы их определения».

## **4.2 Обоснование мероприятий по предотвращению ЧС и разработка порядка действия в случае возникновения ЧС.**

Превентивными мерами по предупреждению пожара могут служить системы звукового и визуального оповещения об опасности, наличие средств пожаротушения и информационных досок с планами эвакуации.

В случае угрозы возникновения ЧС необходимо отключить электропитание, вызвать по телефону пожарную команду, эвакуировать людей из помещения согласно плану эвакуации. При наличии небольшого очага пламени можно воспользоваться подручными средствами с целью прекращения доступа воздуха к объекту возгорания. В качестве подручных средств можно использовать углекислотные огнетушители ОУ-5 высокого давления с зарядом жидкой двуокиси углерода (по ГОСТ 8050-85).

### **Выводы**

Используя все приведенные правила и нормы, касающиеся работы с ПЭВМ, исследование и дальнейшая эксплуатация разрабатываемого модуля не приведет к ухудшению здоровья работника и окружающей его среды.

## **Заключение**

В результате выполнения выпускной квалификационной работы была разработана первая рабочая версия модуля для парсинга, работающая с торговой площадкой Amazon. Дальнейшее развитие работы будет направлено на работу с другими торговыми площадками, а также на увеличение отказоустойчивости в случае неудачи подключения к сайтам торговых площадок, будет проведена оптимизация исходного кода.

Разработан модуль с помощью языка программирования PHP, основой работы парсера модуля является библиотека QueryList, база данных создана с помощью средств СУБД MySQL.

Данный проект позволяет разработчикам систем интеллектуального анализа данных использовать его для своих целей и при необходимости легко модифицировать и расширять. Также за счет своей гибкости модуль легко можно применять и в других системах, где требуется получить данные с сайта в структурированном виде.



## Список литературы

1. Официальный сайт «QueryList». [Электронный ресурс]. URL: <https://querylist.cc/> – Свободный доступ. (дата посещения 21.03.2019).
2. Виталий Г. Блоги и RSS: Интернет-технологии нового поколения [Текст] /Петербург.: «БХВ», 206. – 274.
3. С.Макконнелл. Совершенный код. Мастер-класс / Пер. с англ. – М.: Издательство «Русская редакция», 2017. – 896 стр.: ил.
4. А. Ахо, Дж. Ульман. Теория синтаксического анализа, перевода и компиляции. Т. 1. Пер. с англ. В.Н. Агафонова под ред. В. М. Курочкина. М.: Мир, 1978. 614 с.
5. А. Ахо, Дж. Ульман. Теория синтаксического анализа, перевода и компиляции. Т. 2. Пер. с англ. А.Н. Бирюкова и В.А. Серебрякова под ред. В. М. Курочкина. М.: Мир, 1978. 487 с.
6. Php.net [Электронный ресурс]: Original MySQL API. URL: <http://php.net/manual/en/book.mysql.php>. – Свободный доступ. (дата обращения: 08.04.2019).
7. Чудинов, Игорь Леонидович. Базы данных: учебное пособие / И. Л. Чудинов, В. В. Осипова. Томск: Изд-во ТПУ, 2011. — 140 с.
8. MySQL [Электронный ресурс]: MySQL Documentation. URL: <http://dev.mysql.com/doc/>. – Свободный доступ. (дата обращения: 08.04.2019).
9. Осваиваем парсинг сайта: короткий tutorial на Python [Электронный ресурс]: proglib. URL: <https://proglib.io/p/Web-scraping/>. – Свободный доступ. (дата обращения: 08.04.2019).
10. Простым языком об HTTP [Электронный ресурс]: Хабрахабр. URL: <https://habr.com/ru/post/215117/>. – Свободный доступ. (дата обращения: 09.04.2019).

11. REST API Best Practices [Электронный ресурс]: Хабрахабр. URL: <https://habr.com/ru/post/351890/>. – Свободный доступ. (дата обращения: 13.04.2019).
12. Коды состояний HTTP [Электронный ресурс]: Руководство по REST API. URL: <http://www.restapitutorial.ru/httpstatuscodes.html//>. – Свободный доступ. (дата обращения: 16.04.2019).
13. СанПиН 2.2.2/2.4.1340–03. Санитарно-эпидемиологические правила и нормативы «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы».
14. ГОСТ 12.1.045–84 ССБТ. Электростатические поля. Допустимые уровни на рабочих местах и требования к проведению контроля.
15. СН 2.2.4/2.1.8.562-96. Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки. Санитарные нормы
16. СанПиН 2.2.2.548-96. Гигиенические требования к микроклимату производственных помещений.
17. ГОСТ 12.1.005-88 ССБТ. Воздух рабочей зоны. Общие санитарно-гигиенические требования.
18. СанПиН 2.2.1/2.1.1.1278–03. Гигиенические требования к естественному, искусственному и совмещённому освещению жилых и общественных зданий.
19. ГОСТ 12.1.044-89 ССБТ «Система стандартов безопасности труда. Пожаровзрывоопасность веществ и материалов. Номенклатура показателей и методы их определения».